

Tarea Programada #2

- La tarea se debe entregar en el Github según el profesor haya asignado los grupos.
- Los documentos deben ser entregados en el Dropbox según el profesor haya asignado los archivos.
- Toda tarea debe ser defendida ante el profesor, de tal manera todos los estudiantes deben poder explicar la solución satisfactoriamente.
- ¡Buena Suerte!

UN Visualization

La tarea consiste en 3 partes básicas.

1. Un Web Crawler que inserte datos en un almacenamiento que pueda ser leído por Hadoop.
2. Jobs de Hadoop que tomen la información del Web Crawler, la agregue y además, inserte en una Base de Datos de MariaDB/MySQL
3. Una aplicación Web que lea de la base de datos de MariaDB y retorne los resultados.

Web Crawler

El Web Crawler debe de ingresar a la página <http://data.un.org/Explorer.aspx> y bajar las siguientes estadísticas de la página:

- Estadísticas de Homicidios, los conteos totales y también los que son por sexo.
- La información del a Organización Mundial de la Salud, específicamente la parte de Estadísticas de Salud
- La información de Turismo, todos los datos que están en la página.

Tomen en cuenta que el proceso de procesar e insertar estos datos puede tomar tiempo así que no lo dejen para el último minuto. Esto debe ser la primer parte de la tarea.

El formato específico del archivo en Hadoop lo pueden diseñar ustedes.

Pueden omitir cualquier información que falta o que no se ajuste a los formatos válidos.

Hadoop

Con esta información deben calcular lo siguiente usando Map Reduce y Hadoop:

1. Promedio de Victimas de Homicidio por país
2. Promedio, máximo y mínimo de Homicidios por Región por año
3. Promedio, máximo y mínimo de Homicidios por Subregión por año
4. Promedio de Victimas de Homicidio por país por sexo
5. Promedio, máximo y mínimo de Homicidios por Región y sexo por año
6. Promedio, máximo y mínimo de Homicidios por Subregión y sexo por año
7. Promedio, máximo y mínimo de Homicidios por Región y sexo por década
8. Promedio, máximo y mínimo de Homicidios por Subregión y sexo por década
9. Calcular para país y año cual es el porcentaje de muertes por de personas entre 30 y 70 para
 - a. Cáncer en comparación con el total
 - b. Enfermedades cardiovasculares en comparación al total
 - c. Enfermedades respiratorias en comparación al total
10. Promedio, máximo y mínimo de muertes por país de
 - a. Causas comunicadas
 - b. Causas sin comunicar

c. Heridas

11. La diferencia entre el primer año de medido de edad promedio por país y el último año medido
12. La edad promedio por país por década
13. La cantidad de población por década
14. El porcentaje de fertilidad en adolescentes por cada 1000, por década y país.
15. El porcentaje de fertilidad en adolescentes por cada 1000, por el quintil económico, país y década.
16. El porcentaje de fatalidad infantil antes del primer año por cada 1000, por década y país.
17. El porcentaje de fatalidad infantil antes del primer año por cada 1000, por quintil económico, país y década.
18. El porcentaje de fertilidad por década por país.
19. Total, gastado en salud por país y década
20. Porcentaje gastado en salud por país y década del total del presupuesto del gobierno.
21. Expectativa de vida al nacimiento por país y década.
22. Total, de turistas que entran por país por década.
23. Total, de turistas que salen del país por década

A parte, de estos 23 puntos, deben escoger 5 más a su propio criterio de los mismos datos y calcularlos también.

Esta información debe de insertarse en una base de datos de MariaDB, el diseño de la base de datos lo deben de hacer ustedes.

UI

Se debe crear una página web, en donde se pueda visualizar la información que es calculada por el Hadoop. En general se necesita un gráfico por cada cálculo, en el cual se pueda filtrar por las dimensiones, como país, década, año, etc.

Noten que siempre necesitan algunas dimensiones que no se filtran porque los gráficos las necesitan, pero las demás deben poderse filtrar, por ejemplo, en un gráfico de fertilidad por década por país, pueden dejar la dimensión de fertilidad y década en el gráfico y solo filtrar por país.

Pueden usar una librería para hacer los gráficos como d3.js o chart.js, hay muchas pero les recomiendo usar alguna de esas dos por simplicidad.